A Modified Extended UNIQUAC Model for Proteins

João A.P. Coutinho[a], Isabel M. Marrucho[a], Fernando L.P. Pessoa[b]

[a] CICECO-Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal
[b]Departamento de Eng. Química, Escola de Química, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21949-900 Rio de Janeiro, Brazil

Keywords: Model; Activity coefficient; Protein

Abstract

A modification of the Extended UNIQUAC model is proposed for the description of the non-ideality of protein solutions. Here the Staverman-Guggenheim combinatorial contribution used in Extended UNIQUAC is replaced by the Flory-Huggins term to take into account the size differences between the protein and solvent molecules. This new model allows an excellent description of the activity coefficients in protein systems, for a large range of pH and ionic strengths, with a reduced number of parameters.

Introduction

With the development of biotechnology the interest in production of proteins by microorganisms is quickly increasing. Proteins obtained by fermentation are produced in complex broths containing inorganic salts, sugars, organic acids and cells from where the recovery of the higher value compounds may prove difficult. A number of methods for protein purification, applicable at bench scale, have been developed (Scopes, 1994). However the scale-up and optimisation of these techniques at an industrial level it is only possible if a mathematical description of the process is available. For that purpose the thermodynamic description of the protein non-ideality in the broth and its dependence on the pH, ionic strength, temperature and on the presence of secondary compounds such as sugars (simple or polysaccharides) and polymers is essential.

A number of attempts to describe the behaviour of proteins in solution have been reported in the literature. Most approaches are based on the Potential of Mean Force (Haynes et al., 1992; Kuehner et al., 1996; Tavares and Sandler, 1997) or on equations of state based on hard sphere interaction potentials (Hino and Prausnitz, 1999). Excess Gibbs energy models have seldom been applied to the description of protein systems. Engineering local composition models such as UNIQUAC, UNIFAC or NRTL although very used for amino acid systems (Pinho et al., 1994; Gude et al., 1996; Khoshkbarchi and Vera, 1998; Bellot et al., 1999; Soto et al., 1999) with some success, were applied to proteins only by Agena et al. (Agena et al., 1997 and 1999) in collaboration with one of the authors. Agena's work it is, however, a very crude approach to the modelling of proteins in solution as a conventional UNIQUAC model is used, not taking into account the large size differences between the protein and solvents, neither considering the long range forces that arise from the electrostatic interactions between the protein and the solvent.

On this work it will be shown that a new g^E model, developed by modifying the Extended UNIQUAC model, can provide an excellent description of the activity coefficients of the protein in solution with a reduced number of parameters and that these parameters can be used to predict the behaviour of the protein at pH's and ionic strengths other than those used for the parameter estimation. The protein activity coefficients derived from the osmotic pressure data reported by Haynes et al. [2] for the α-Chymotrypsin are used to validate the proposed model.

Model

The Extended UNIQUAC model has been proposed for the description of classical electrolyte systems, i.e. aqueous systems of small (~4 Å), inorganic ions with a

charge of 1 or 2 such as Na^+, K^+ Cl^- or SO_4^{2-} (Nicolaisen et al., 1993; Thomsen et al., 1996). It has never been applied to organometallic electrolytes neither to any sort of organic electrolytes. Nevertheless the model proved to be very successful in the description of SLE and VLE for inorganic systems, as well as for a number of other properties such as osmotic coefficients, heat capacities and heats of solution. For this reason it was chosen among the different electrolyte models available for the description of behaviour of proteins in solution.

Proteins differ from the small inorganic ions in size (typically larger than 40 Å with molecular weights higher than 20000), in complexity (they possess a large and diversified number of functional groups and elaborate structure), and in charge (proteins can change from a net charge of +60 to -60 within a few pH units). For such system the excess Gibbs energy will have three contributions

$$G^E = G^E_{Combinatorial} + G^E_{Residual} + G^E_{Debye-Huckel} \quad (1)$$

The combinatorial term will account for the entropic interactions arising from size and shape differences between the molecules. The Extended UNIQUAC uses the Staverman-Guggenheim term to represent these interactions. However it is known that for very asymmetric systems the Staverman-Guggenheim does not produce a good description of the non-ideality of the systems (Kikic et al., 1980; Kontogeorgis et al., 1993) and it was replaced by the Flory-Huggins term to better take into account the very large size differences between the protein, the water and the other ions. As shown in Figure 1, although for protein concentrations higher than 1 mol% the two terms are virtually identical, for dilute solutions the Staverman-Guggenheim shows a very implausible behaviour. The Flory-Huggins combinatorial term here used is given by:

$$\frac{G^E_{Combinatorial}}{RT} = \sum_i x_i \ln\left(\frac{f_i}{x_i}\right) \quad (2)$$

The combinatorial-free volume terms usually used for polymer solutions (Kontogeorgis et al., 1993) were not adopted since the water has a free volume similar to polymers (Rasmussen and Rasmussen, 1989) and thus it is not necessary to use a free volume contribution for polymer aqueous systems.

The UNIQUAC residual term that accounts for the energetic short-range interactions is

$$\frac{G^E_{Residual}}{RT} = -\sum_i x_i q_i \ln\left(\sum_k \theta_i \psi_{ki}\right) \quad (3)$$

where q_i are the UNIQUAC surface area parameters, θ_i the surface area fractions and the parameter ψ_{kl} is given by

$$\psi_{kl} = \exp\left(-\frac{u_{kl} - u_{ll}}{T}\right) \quad (4)$$

with the UNIQUAC interaction parameters u_{kl}=u_{lk}. The surface area parameter for the protein q_p=700.2723 was obtained from a correlation by Agena et al. (1997). The parameters for water (q_w=1.400) were obtained from Thomsen et al. (1996).

Despite the change in the combinatorial term it was assumed that the UNIQUAC parameters available in the Extended UNIQUAC parameter table for the ion-ion, ion-water and water-water interactions can be used without reestimation. This was found acceptable in previous works where UNIQUAC or UNIFAC interaction parameters fitted to small molecules with a given combinatorial term are used for polymer systems with a new combinatorial term (Kontogeorgis et al., 1993; Coutinho and Stenby, 1996).

The reason for this can be understood from Figure 1. For concentrations higher than 1 mol % the various combinatorial terms are essentially identical. Since for simple systems most of the data available belongs to the region where the two combinatorial terms are identical, interaction parameters adjusted for the Flory-Huggins would be similar to the parameters obtained with the Staverman-Guggenheim combinatorial term. This approach allows only a small number of interaction parameters to be fitted for new systems.

The Debye-Huckel term for the long range electrostatic interactions is

$$\frac{G^E_{Debye-Huckel}}{RT} = -x_w M_w \frac{4A}{b^3}\left[\ln\left(1+b\sqrt{I}\right) - b\sqrt{I} + \frac{b^2 I}{2}\right] \quad (5)$$

where x_w is the mol fraction and M_w the molar mass of water. A is a constant that in the 273.15 to 373.15 K range can be approximated by (Thomsen et al., 1996)

$$A = 1.131 - 1.335\times10^{-3}(T-273.15) + 1.164\times10^{-5}(T-273.15)^2 \quad (6)$$

and b is a constant that depends on the size of the ions. For inorganic ions (~4 Å) it can be taken as 1.5 (kg mol^{-1})$^{1/2}$. For the protein, with a size of ~40 Å, b was taken as 15 (kg mol^{-1})$^{1/2}$. I is the ionic strength given by

$$I = \frac{1}{2}\sum_i m_i z_i^2 \quad (7)$$

where m_i is the molality of ion i and z_i its charge.

Unlike the charge of simple ions, the protein net charge and its dependence on the pH is not easy to obtain, either experimentally or by calculation, with accuracy. The approach used in this work was to treat the protein net charge ($z_{protein}$) as a fitting parameter. A discussion on the physical sense of the fitted protein charges is presented in the next section.

The activity coefficient for species i, γ_i is obtained by partial molar differentiation of the excess Gibbs energy according to

$$\ln g_i = \left[\frac{\partial\left(\frac{nG^E}{RT}\right)}{\partial n_i}\right]_{P,T,n_{j\neq i}} \quad (8)$$

As usual with electrolytes the asymmetrical convention is adopted in this work.

Results

Using the model proposed above it was attempted to model the activity coefficients for aqueous solutions of α-Chymotrypsin obtained from osmotic pressure measurements (Haynes et al., 1992) to evaluate its capacity. The asymmetric molal activity coefficients are calculated from the osmotic pressures following the approach of Wills et al. (1993). From the osmotic pressure data, Π, over the molar protein concentration, c_p, it was possible to obtain the virial coefficients, B_k, using the equation

$$\frac{\Pi}{RT} = c_p\left(1 + B_2 c_p + B_3 c_P^2 + ...\right) \quad (9)$$

Solute molal activity coefficients dependence with the molal composition, m, are given by

$$\ln \gamma_p = 2C_2 m_p + \frac{3}{2} C_3 m_p^2 \quad (10)$$

The coefficients, C_k, are related to the virial coefficients, B_k, by

$$C_2 = \left(B_2 - \nu_p M_p\right)\rho_s \quad (11)$$

$$C_3 = \left(B_3 - 2B_2 \nu_p M_p + \left(\nu_p M_p\right)^2\right)\rho_s^2 \quad (12)$$

where ν_p and M_p represent the partial specific volume and the molecular weight of the protein with values of 0.736 cm3/g (Agena et al., 1997) and 25651 g/mol (Berman et al., 2000), and ρ is the solvent density.

The data used have protein concentrations ranging from 0 to 9 g/L and pH's ranging from 3 to 12 in 0.1 M Potassium Sulphate buffer (I=0.3 M) except for two cases for which the buffer ionic strength, at pH=3, was 0.03 and 0.15 M. Information about the α-Chymotrypsin composition and structure were obtained from Berman et al. (2000). The ionisable groups present are reported in Table 1.

The pH dependency of the protein activity coefficient

The pH dependency of the protein solubility is well known and can be found in any biochemistry textbook (Voet and Voet, 1995). It follows a U shaped curve with a minimum solubility at the isoelectric point that increases for both higher and lower pH's. The asymmetrical molal activity coefficients obtained from osmotic pressure measurements also show this behaviour. The activity coefficients have values close to 1 near the isoelectric point decreasing with pH to both sides of the isoelectric point.

As discussed above the interaction parameters, u_{kl}, for ions and water already available for the Extended UNIQUAC (Thomsen et al., 1996) were used on this work without reestimation to minimize the number of parameters to fit. A parameter table for the Extended UNIQUAC can be found in the work by Thomsen et al. (1996). To prevent interferences of the electrostatic contribution on the optimised parameter values, the UNIQUAC interaction parameters for the protein (protein-protein, protein-water, protein-K^+ and protein-SO_4^{2-}) were fitted to the data available at isoelectric point (pH=8.25). The parameters obtained are reported in Table 2. Their values indicate that the interaction between the protein and ions are negligible. Further work is still required to support this result. Since the short-range interactions described by the residual term will remain the same over the entire range of pH the set of parameters fitted at the isoelectric point was used to describe the energetic interactions for all pH values.

The long-range interactions generated by the protein net charge and their dependence with the pH are described by the Debye-Huckel term. As mentioned before the protein net charge is not easily available with the exception of the isoelectric point and the very high and low pH region. For this reason it was decided to fit the protein charge at each pH value. The fitted values are presented in Figure 2. A comparison between the experimental and model results for the activity coefficients reported in Figures 3a and 3b shows the adequacy of the model for the description of the non-ideality of protein solutions. The proposed model, compared to previous works (Agena et al., 1997 and 1999), uses only a reduced number of parameters to describe the activity coefficient data. These are the UNIQUAC interaction parameters reported on Table 2 plus the protein net charge fitted at each pH value.

Since the UNIQUAC interaction parameters are pH independent the (combinatorial plus residual) term contributions to the activity coefficient are always identical to the values at the isoelectric point (pH=8.25). The main contribution to the

activity coefficient at pH's removed from the isoelectric point is the Debye-Huckel term, as can be seen in Figures 3a and 3b comparing the activity coefficients at different pH values. This importance of the Debye-Huckel term is due to the large protein net charge.

The fitted values of the net protein charge are compared in Figure 2 with the values calculated assuming that the pKa's of the amino acids on the protein are the same as the free amino acids. This approach is not exact. The pKa's for amino acids in proteins may differ from their free form values by as much as 3 pH units (Haynie, 1999) and they are also dependent on the salts present in solution (Voet and Voet, 1995). Nevertheless this approach provides a fair and simple estimate of the protein net charge dependence with pH. The comparison between the fitted and the calculated values shows an excellent agreement between the two curves except for pH values between 4 and 7. This is probably due to the position of some acidic amino acids in the protein structure that are less available to the solvent and consequently more difficult to ionise.

The results presented show that the proposed model can be used to describe the activity coefficients for proteins if experimental data at the isoelectric point is available to fit the short-range interaction parameters u_{kl}. If data at other pH values is available the model can also be used to describe the pH dependence of the activity coefficients by fitting the protein net charge. Yet if this data is not available the protein net charge assessed from the pKa's of the free amino-acids can be used to predict the activity coefficients for the protein at different pH values with a reasonable degree of confidence. This is particularly valid for the extreme values of pH and around the isoelectric point where the calculated protein net charge is less prone to be affected by amino-acids with pKa's substantially different from their free-form values due to their location within the protein.

Ionic strength dependency of the protein activity coefficient

In the same way as for the pH, the ionic strength dependence of the protein activity coefficients is well established (Voet and Voet, 1995). For very low salt concentrations the protein solubility increases with increased salt content creating the *salting-in* region while for higher salt concentrations the increase in salt content promotes the precipitation of the protein in a phenomenon known as *salting-out*. This originates the typical dependency of the protein solubility curves with ionic strength with the shape of a reverse U. In terms of activity coefficients a decrease is found in the salting-in region followed by an increase in the region of salting-out.

Predictions using the net charge and interaction parameters previously estimated were done for the effect of the ionic strength on the solubility of the protein at pH =3. Results shown in Figure 4 indicate that experimentally there is a decrease in the activity coefficients between the ionic strengths of 0.03 and 0.15 M corresponding to a salting-in region followed by an increase between 0.15 and 0.3 M characteristic of a salting-out region. The model however describes a continuous decrease of the activity coefficients as the ionic strength is reduced from 0.3 to 0.03 M. The 0.3 and 0.15 M systems are well described. It must be emphasized that the behaviour at 0.15 M is a pure prediction. Yet the model cannot describe the salting-in region and predicts activity coefficients for ionic strengths of 0.03 M much lower than the experimental values. This limitation may be related to the protein-ion interactions that were taken as negligible as discussed above. Studies to improve the model for simultaneous description of the salting-in and salting-out regions are being undertaken.

Temperature dependency of the protein activity coefficient

Proteins may present very different temperature dependence solubilities. Cristopher et al. (1998) present results for a number of proteins with temperature dependent solubilities including a number of proteins with retrograde solubility (the α-Chymotrypsinogen may have such behaviour) and some proteins which solubility does not seems to be significantly affected by the temperature. Although the data for the system under study is available only at 25 ºC and the temperature dependence of the protein behaviour in solution cannot be investigated or correlated, the temperature dependence of the Extended UNIQUAC parameters

$$u_{ij} = u_{ij}^{o} + u_{ij}^{t}\left(T - 298.15\right) \qquad (13)$$

should be able to describe any temperature dependence of the activity coefficients within a reasonable temperature range.

Conclusions

A modified version of the Extended UNIQUAC model is proposed for proteins and other polyelectrolytes in aqueous solution. The activity coefficients for α-Chymotrypsin obtained from osmotic pressure measurements in a wide range of pH and ionic strength were used to validate the proposed model. The interaction parameters for water and ions used were taken from Extended UNIQUAC tables and the protein interactions were fitted to data at the isoelectric point. The protein net charge was used as a fitting pH dependent parameter but it was shown to be in close agreement to the calculated protein net charge. Due to the large protein net charge the Debye-Huckel is the key term in the activity coefficient model. A good description of the activity coefficients is achieved over the entire pH region studied and the parameters estimated allow for predictions of the protein behaviour on the salting out region.

Aknowledgements

This work was developed in the framework of a project financed by ICCTI (Portugal) and CAPES (Brasil). Thanks are due to Prof. Carlos M. Silva (UA) and Prof. Maria Alice Z. Coelho (UFRJ) for fruitful discussions.

References

S.M. Agena, D.L. Bogle, F.L. Pessoa, Biotech. Bioeng., 55 (1997) 65-71.
S.M. Agena, M.L. Pusey, D.L. Bogle, Biotech. Bioeng., 64 (1999) 44-50.
J.C. Bellot, R.V. Tarantino, J.S. Condoret, AIChE J., 45 (1999) 1329-1341.
H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne, *Nucleic Acids Research*, 28 (2000) 235-242.
G.K. Cristopher, A.G. Phipps and R.J. Gray, J Crystal Growth, 191 (1998) 820-826.
J.A.P. Coutinho and E.H. Stenby, Ind. Eng. Chem. Res., 35 (1996) 918-925.
M.T. Gude, L.A.M. van der Wielen, K.Ch.A.M. Luyben, Fluid Phase Equil., 116 (1996) 110-117.
C.A. Haynes, K. Tamura, H.R. Korfer, H.W. Blanch and J.M. Prausnitz, J. Phys. Chem., 96 (1992) 905-912.

D.T. Haynie, Biological Thermodynamics, Cambridge University Press, Cambridge, 1999.
T. Hino, J.M. Prausnitz, AIChE J., 45 (1999) 622-632.
M.K. Khoshkbarchi, J.H. Vera, Ind. Eng. Chem Res., 37 (1998) 3052-3057.
I. Kikic, P. Alessi, P. Rasmussen and Aa. Fredenslund, Can. J. Chem. Eng. 58 (1980) 253-7.
G.M. Kontogeorgis, Aa. Fredenslund and D. Tassios, Ind. Eng. Chem. Res., 32 (1993) 362-372.
D.E. Kuehner, H.W. Blanch, J.M. Prausnitz, Fluid Phase Equil., 116 (1996) 140-147.
H. Nicolaisen, P. Rasmussen and J.M. Sørensen, Chem. Eng. Sci., 48 (1993) 3149-3158.
S.P. Pinho, C.M. Silva and E.A. Macedo, Ind. Eng. Chem Res., 33 (1994) 1341-1347.
D. Rasmussen and P Rasmussen, Chem. Eng. Prog., February 1989, 50-56.
A. Soto, A. Arce, M.K. Khoshkbarchi, J.H. Vera, Fluid Phase Equil., 158-160 (1999) 893-901.
R.K. Scopes, Protein Purification: Principles and Practice, 3rd ed., Springer Verlag, 1994.
F.W. Tavares, S.I. Sandler, AIChE J., 43 (1997) 218-231.
K. Thomsen, P. Rasmussen and R. Gani, Chem. Eng. Sci., 51 (1996) 3675-3683.
D. Voet and J.G. Voet, Biochemistry, 2nd ed, John Wiley & Sons, 1995.
P.R. Wills, W.D. Comper, J.D. Winzor, Arch. Biochem. Biophys., 300 (1993) 206-212.

Table 1- Acid and basic groups on α-Chymotrypsin (Berman et al., 2000)

Basic groups		**Acid groups**	
Arginine	4	*Tyrosine*	4
Histidine	2	*Cysteine*	10
Lysine	14	*Ac. Glutamic*	5
NH2 terminal	1	*Ac Aspartic*	9
		COOH terminal	1

Table 2- UNIQUAC interaction parameters, u_{kl}, for α-Chymotrypsin estimated using data at the isoelectric point.

Interaction parameter	**Parameter value [K]**
$u_{\alpha\text{-Chymotrypsin}/\alpha\text{-Chymotrypsin}}$	-348.96
$u_{\alpha\text{-Chymotrypsin/Water}}$	-56.1572
$u_{\alpha\text{-Chymotrypsin}/K^{+}}$	5311
$u_{\alpha\text{-Chymotrypsin}/SO4^{2-}}$	3011

Figure Captions

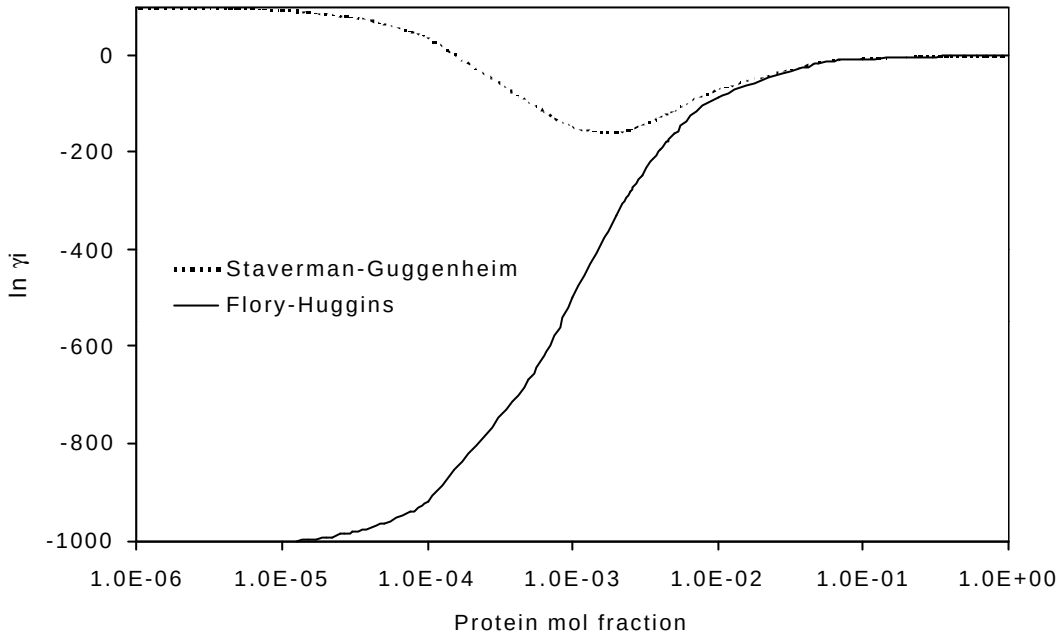

Figure 1- Comparison between the Staverman-Guggenheim and Flory-Huggins combinatorial terms for α-Chymotrypsin in aqueous solution.

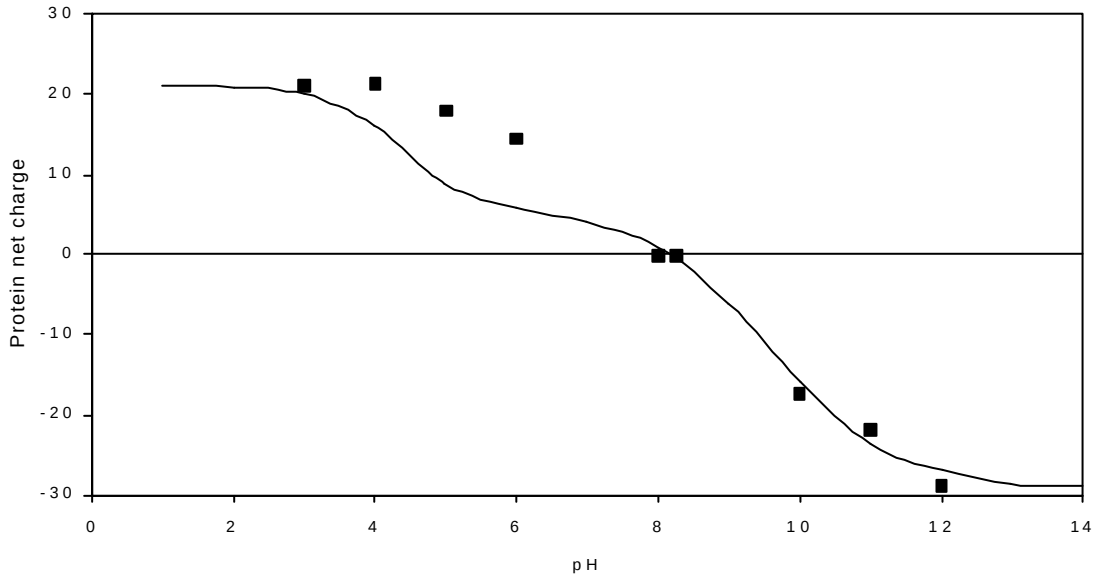

Figure 2- Comparison between the fitted and estimated protein net charge pH dependency.

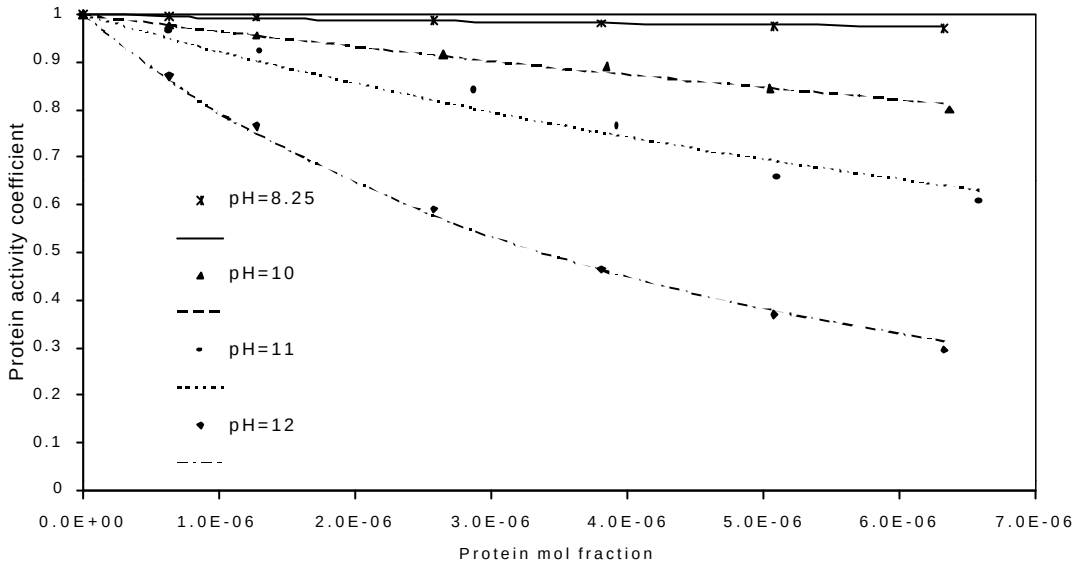

Figure 3a- Correlation of the activity coefficients with Extended UNIQUAC for pH's above pI.

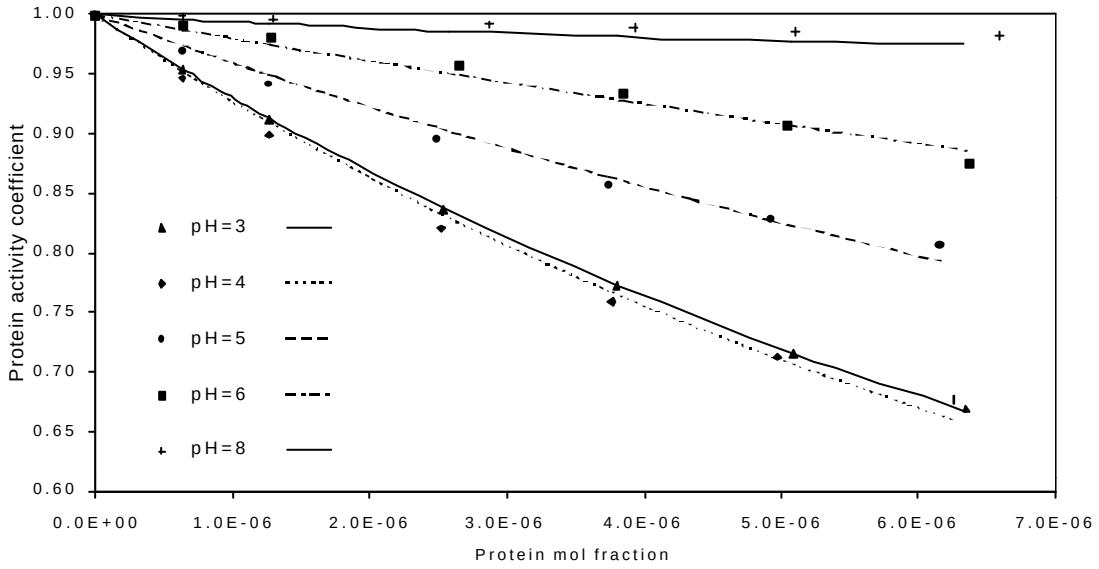

Figure 3b- Correlation of the activity coefficients with Extended UNIQUAC for pH's below pI.

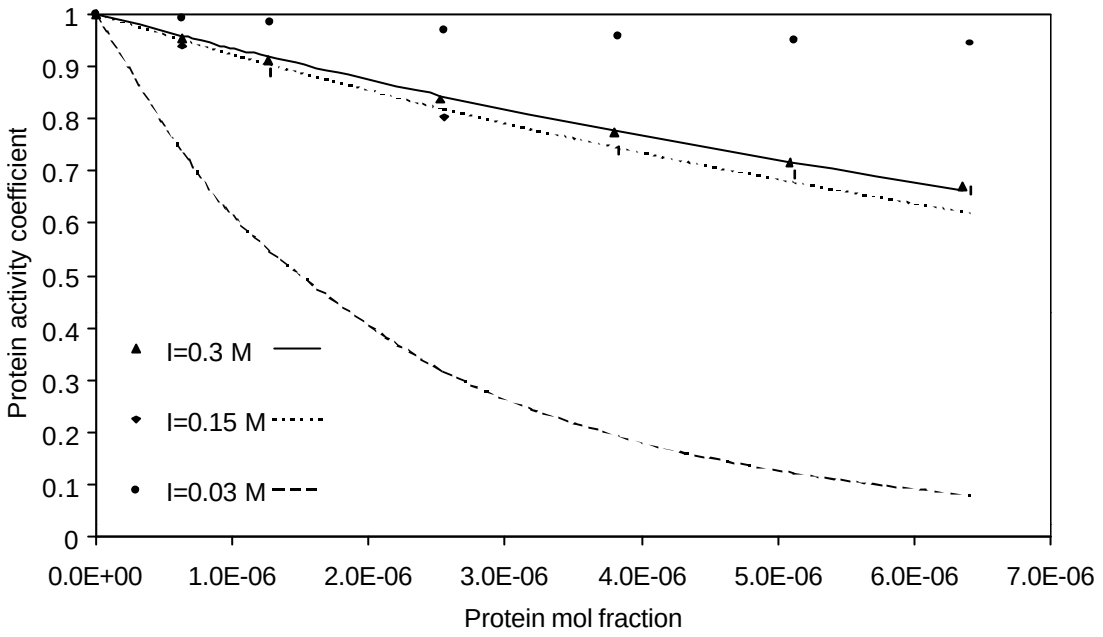

Figure 4- Experimental and predicted ionic strength dependence of the activity coefficients.